

Laboratorio Colisiones

a) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?

Implementaría 6.

Artists (una arraylist con los nombres de los artistas), ArtistsID(un mapa con key Constituent ID y de valor tiene la info del artista). Lo mismo con piezas, es decir, piezas y piezasID. Estos son un array y un mapa respectivamente. El mapa tiene como key ConstituentID de los artistas y como valor la info del piece.

Por otro lado, se encuentran los índices de nacionalidad y médium que son mapas con Chaining. Esta configuración permite una mayor eficiencia que en la anterior (donde solo había un índice en artistas, otro en piezas, médium y el cuarto era nationality). Es más eficiente porque al tener los constituent ID disponibles en el key de artistas y piezas ya no se requiere hacer dos ciclos.

b) Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?

Realmente debería funcionar con ambos tipos de manejo de colisiones (tanto con probing como con separate chaining) lo que realmente los diferencia es que separate chaining lidia con el gasto de memoria puesto que no deja espacios sin usar. Durante el linear probing al final del proceso quedan slots sin utilizar, lo que puede ser perjudicial para el uso eficiente de memoria.

c) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?

Para linear probing, al estar trabajando con un archivo grande se llena constantemente los slots permitidos para el factor de carga (que se asigna entre 0.5 y 0.7 para evitar colisiones). Y para chaining debe ser mayor claramente a 1, de lo contrario cada slot solo podría contener a un elemento (llave-valor)

d) ¿Qué diferencias en el tiempo de ejecución notan al ejecutar la carga de los datos al cambiar la configuración de Linear Probing a Separate Chaining?

Para responder a esta pregunta hicimos pruebas usando las diferentes configuraciones de chaining y probing y las comparamos para evidenciar cual retornaba mejores resultados de tiempo. Se realizaron 3 pruebas para usar su media como tiempo promedio de ejecución.

(Chaining, loadfactor= 4.00): 6.591, 6.363, 6.416

(Chaining, loadfactor= 2.00): 6.333, 6.549, 6.90

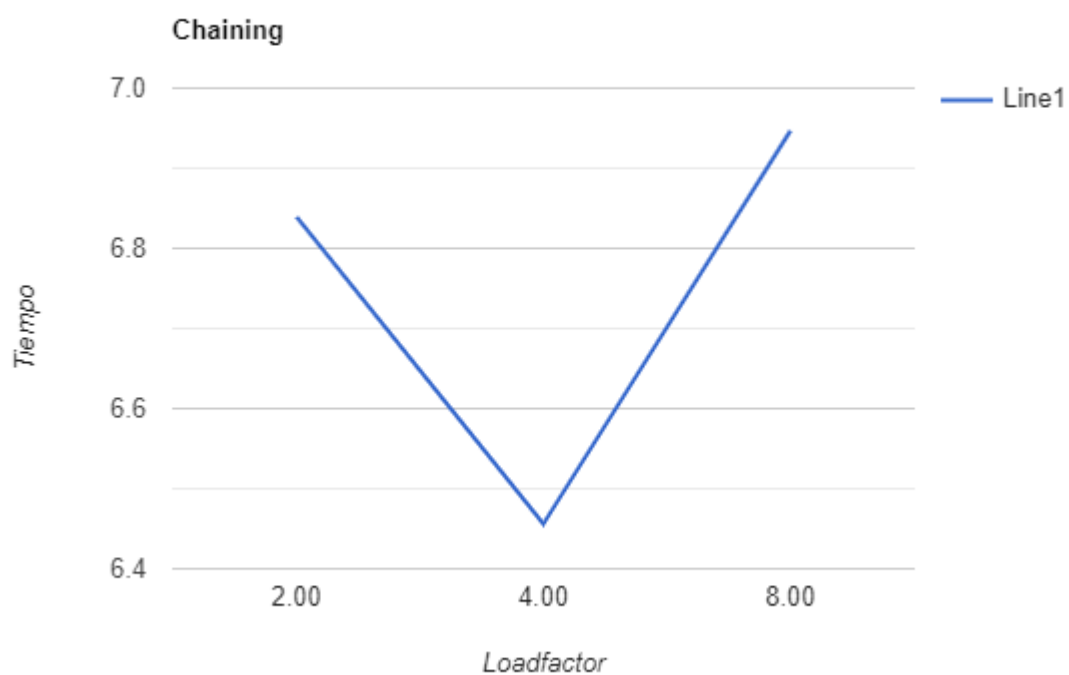
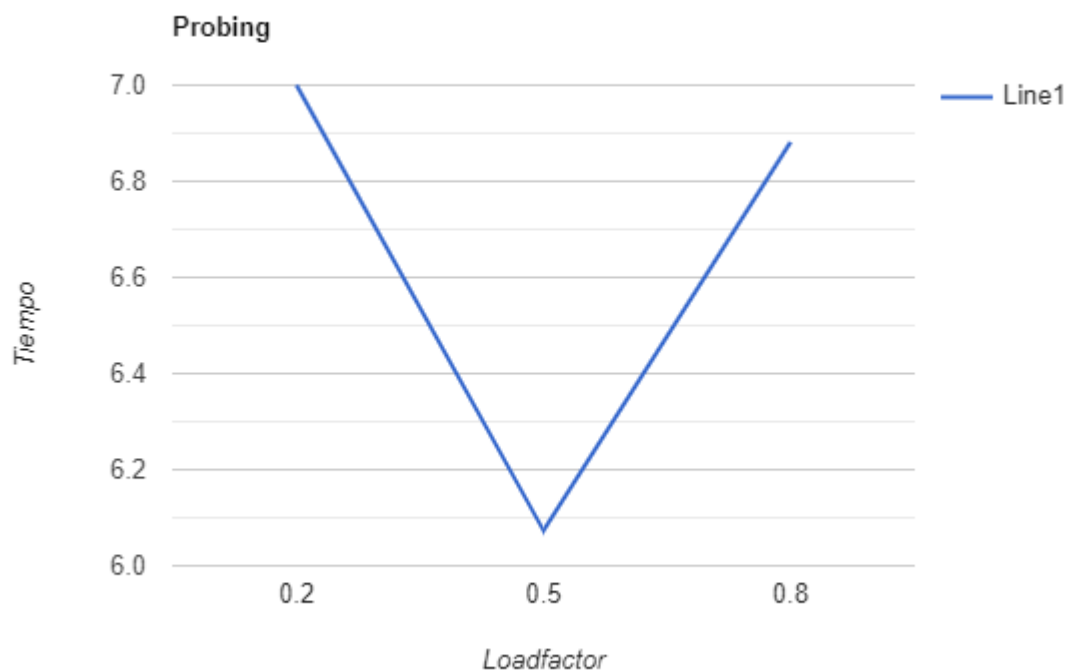
(Chaining, loadfactor= 8.00): 7.162, 7.052, 6.63

(Probing, loadfactor= 0.5): 6.074, 6.11, 6.038

(Probing, loadfactor= 0.2): 7.122, 6.595, 7.283

(Probing, loadfactor= 0.8): 7.095, 6.79, 6.760

de lo anterior obtuvimos las siguientes gráficas:



Notamos que el tiempo de carga es considerablemente mayor para 2.00 y 8.00 en chaining y análogamente en 0.2 y 0.8 en probing. En el caso de probing cuando el factor de carga es muy pequeño el map debe incrementar su tamaño muy seguido ya que alcance el límite muy fácilmente. Por otro lado, cuando el factor de carga es muy grande hay mas probabilidades de colisión entre los elementos pues no hay mucho espacio para elementos nuevos. Naturalmente el factor de carga más beneficioso es un punto medio que permite a

la lista introducir muchos elementos sin incrementar los slots. En el caso del chaining podemos notar pocos slots posibles si hay colisión resultan en un mayor tiempo de ejecución, lo que va en contra de la teoría ya que se supone que entre más nos acerquemos a 1 mejor será la velocidad del programa con chaining. Por otro lado, cuando hay muchos slots disponibles el tiempo se reduce porque tiene que recorrer con complejidad n en cada uno, si n es más grande la complejidad aumenta.

e) ¿Qué configuración de ADT Map escogería para el índice de técnicas o medios?, especifique el mecanismo de colisión, el factor de carga y el número inicial de elementos. Para el índice de medio escogería probing con un factor de carga 0.5 ya que según las pruebas es la configuración con mejores resultados. Además, contiene 21250 por lo que es considerablemente grande y requiere amplio espacio para que no haya muchas colisiones. El número inicial de elementos sería el primo más cercano al doble de 21250

f) ¿Qué configuración de ADT Map escogería para el índice de nacionalidades?, especifique el mecanismo de colisión, el factor de carga y el número inicial de elementos. En este caso como hay pocas nacionalidades escogería chaining ya que podría funcionar sin colisiones, con un factor de carga de 1.00 se garantiza que solo habrá un elemento en la lista. El número inicial de elementos sería el primo más cercano al doble de 114.